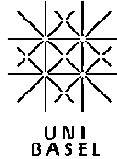


Seminar

Angewandtes Information Retrieval

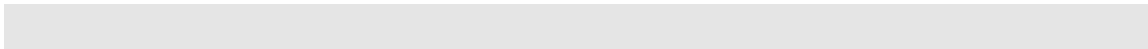
Sommersemester 2007



Philosophisch-Historische Fakultät
Geisteswissenschaftliche Informatik
Universität Basel
Petersgraben 51
CH-4003 Basel, Schweiz
Seminarleiter: Dr. Urs Hengartner

Geschichte und Trends im IR

Edy Portmann
Telefon: ++41-41-2400632
E-Mail: edy_portmann@gmx.ch



Stichworte:

Archie, Deep Web, Geschichte des Information Retrievals, Gopher, Information Retrieval, Internetgeschichte, Invisible Web, Lokalisierte Suche, Mosaic, Personalisierung von Suchmaschinen, Spezialisierte Suchmaschinen, Suchmaschinen-Trends, Trends im Information Retrieval, Veronica, WAIS

Zusammenfassung

Diese Seminararbeit wurde im Rahmen des Seminars Angewandtes Information Retrieval geschrieben und beschäftigt sich mit der Geschichte der Daten, der Internetgeschichte und dem untrennbar dazugehörenden Information Retrieval, welches als Wiedergewinnung von bereits zur Verfügung stehenden Daten gesehen werden kann.

In einem ersten Teil befasst sich diese Arbeit mit der Geschichte der menschlichen Datensammlung und Speicherung. Hier wird die Geschichte von den Anfängen der Datensammlung bis hin zur heutigen digitalen Zeit durchlaufen.

Im zweiten Teil werden die Evolution und Funktionsweise verschiedener Systeme vorgestellt, wobei eine Trennung vorgenommen wird in die Geschichte des Information Retrieval, als Antwort auf die Datenmengen, welche durch die Evolution hervorgebracht wurde, und dann wird auf heutige Trends des Information Retrievals eingegangen. Hierbei werde ich noch einmal die grundlegenden Probleme der Informationssuche aufzeigen und die aktuellen Forschungen in diesem Gebiet erwähnen.

Danach wird ein Fazit gezogen und kritisch Stellung bezüglich der aktuellen Trends eingenommen. In meinem Schlusswort gebe ich meine Vision einer Suchmaschine wieder, wie diese in Zukunft aussehen könnte.

Inhaltsverzeichnis

Stichworte:	II
Zusammenfassung	II
Inhaltsverzeichnis	III
1 Einleitung	1
2 Geschichte der Datenspeicherung	3
2.1 Höhlenmalerei, Bücher und Informatik	3
2.2 50 Jahre Internetgeschichte	4
2.2.1 <i>Die ersten Schritte</i>	4
2.2.2 <i>Die siebziger und achtziger Jahre</i>	5
2.2.3 <i>Die Schritte der Neunziger bis heute</i>	7
2.3 Das Internet heute	8
3 Evolution und Funktionsweise verschiedener Systeme im IR	10
3.1 Ausgewählte Momente in der Geschichte des IR	10
3.1.1 <i>Archie</i>	10
3.1.2 <i>WAIS</i>	10
3.1.3 <i>Gopher</i>	11
3.1.4 <i>Veronica</i>	11
3.1.5 <i>Mosaic</i>	12
3.2 Gegenwärtige Trends im IR	12
3.2.1 <i>Informationsressourcen</i>	13
3.2.2 <i>Spezialisierte Suchmaschinen</i>	13
3.2.3 <i>Personalisierung</i>	14
3.2.4 <i>Lokalisierte Suche</i>	14
4 Fazit / Schlussfolgerungen	16
5 Abbildungsverzeichnis	17
6 Literatur	18
6.1 Bücher	18
6.2 Wissenschaftliche Arbeiten	18
6.3 Internet	18

1 Einleitung

Der Mensch hatte von jeher das Bedürfnis seine Ideen an andere weiter zu geben. Unter dem Mikroskop der Zeit betrachtet möchte ich, auf wenige Zeitpunkte gerafft, hierzu einen kleinen Exkurs in die Geschichte der Menschheit unternehmen. Als Startpunkt und Initialzündung kann man heute aus ferner Sicht die Höhlenmalereien und Steinmeisselungen als erste Technik zur Übermittlung von Wissen an andere benennen. Diese primitiven Techniken veralteten allerdings und wurden glücklicherweise durch die Erfindung der Papyrusrollen abgelöst, welche wiederum durch den mittelalterlichen Buchdruck von Johannes Gutenberg abgelöst werden konnte. Dies ermöglichte die schnelle Verbreitung von Ideen. Die technischen Beschränkungen konnten immer mehr behoben werden und so folgte eine kleine Explosion von Daten schon vor dem digitalen Zeitalter und somit auch vor der Erfindung des Internets. Hingegen entfaltete sich die Datenflut in der heutigen, digitalen Zeit voller Bits und Bytes zu einer unüberschaubaren Masse. Heutzutage kann sich jedermann ohne grossen Aufwand der Welt mitteilen und seine Informationen durch gegebene, einfach zu bedienende, technische Möglichkeiten an andere weitergeben.

Gegenwärtig ist das Internet in aller Munde und es entwickelt sich, erfreulicherweise, immer rasanter weiter. Im Internet wird man fündig, egal was man finden möchte, sofern man es findet! Häufig findet man dabei, analog zum altbekannten Sprichwort; „Vor lauter Bäumen den Wald nicht mehr sehen“, im Internet vor lauter Daten die relevanten Daten nicht mehr. Deshalb drängt sich die Frage auf, ob die Datenexplosion im Internet wirklich nur erfreulich sei.

Wurde also durch das Internet alles besser? Findet man wirklich schneller korrekte und relevante Daten? Gab es eventuell früher bessere Suchmaschinen? Dies sind die Fragen, mit welchen wir uns in der heutigen Zeit beschäftigen sollten. Wichtig ist der Aspekt, dass Quantität nicht mit Qualität verwechselt werden darf. Zudem kann gefragt werden, ob es Möglichkeiten gibt mit den modernen Mitteln der Suchmaschinen der Informationsflut zu trotzen.

Die wohl berühmteste Suchmaschine ist Google, denn nach dem Namen einer Suchmaschine gefragt, bekommt man gegenwärtig ausser Google keine andere Suchmaschine aufgezählt¹. So wurden dazu verschiedene Untersuchungen durchgeführt. Unter anderem wurde nach Angaben von ComScore [Come07] die Webseiten von Google im März 2007 zum ersten Mal von mehr Besuchern² angeklickt als diejenigen vom Rivalen Microsoft. Die verschiedenen Google-Seiten wurden nach Angaben von ComScore von 528 Millionen Usern besucht, was 13 Prozent mehr als im März 2006 sind. Diese Steigerung verdankt Google vor allem dem Erwerb des Videoportals YouTube. Knapp weniger Besucher als Google verbuchte Microsoft auf seinen Auftritten mit nur 527 Millionen Gästen. Auf dem dritten Platz liegt Yahoo mit 473 Millionen Zugriffen, gefolgt von Time Warner mit 272 Millionen und Ebay mit 256 Millionen Besuchern. Wikipedia schaffte es mit 212 Millionen auf den sechsten Rang. Damit hatte die online-Enzyklopädie 82 Prozent mehr Zugriffe als im März 2006.

Als weitere Untersuchung wurde von der Marktforschungsgruppe Millward Brown Optimor [Mill07] ermittelt, dass Google die wertvollste Technologiemarke ist und es Google auch im Jahr 2007 gelang, den Microsoft Konzern vom ersten Platz zu verdrängen. Mit dieser Arbeit möchte ich wie bereits erwähnt, die Geschichte der Informationswiedergewinnung aufarbeiten und zugleich einen kurzen Abstecher in die Forschungstrends von heute unternehmen.

Starten werde ich diese Seminararbeit mit einem Rückblick auf die Geschichte der Menschheit und ihres Drang, Informationen für andere verständlich aufzuarbeiten und darzustellen. Dies führt mich im ersten Teil unter anderem zurück in eine Zeit, als Daten noch in Stein

¹ Vgl. Arbeit von Virginia Siderkeviciute [SIDE05] und zugehörige Literatur.

² Zur besseren Lesbarkeit des Textes wird im ganzen Bericht auf eine geschlechtsspezifische Formulierung verzichtet.

gehauen wurden. Schnell werde ich mich dann aber der heutigen Zeit annähern. Ab 1960 schlag ich ein gemächlicheres Tempo an, denn von da an möchte ich mich mit der Entwicklung des Internets beschäftigen. Ich betrachte das Internet hierzu als Ansammlung verschiedenster Daten, welche mit Suchmaschinen geordnet und durchsucht werden können. Ich werde in meiner Arbeit darum als nächsten Punkt auf die Evolution des Information Retrieval eingehen, welcher sich parallel zur genannten Geschichte entwickelte. Seit ungefähr 4000 Jahren organisiert sich der Mensch Informationen für den späteren Gebrauch und seit diese Informationen auf ein paar Bücher anwuchsen, wurde es nötig nach Informationen suchen zu können. So werde ich den Versuch unternehmen die Trends des heutigen Information Retrieval genauer zu beleuchten.

Ich werde hierzu auf verschiedene Technologien eingehen, empfehle aber all jenen, welche sich für (eine) spezielle Technik(en) interessieren dem unter Punkt „Literatur“ genannten Verweise zu folgen. Da ich in dieser kurzen Aufarbeitung nicht allzu tief auf ein Thema eingehen kann und die technischen Details aussen vor lasse, möchte ich die empfohlene Literatur ans Herz legen, um einen weiterführenden Überblick zu erhalten. Abschliessen werde ich mit einem kritischen Fazit bezüglich der Trends im Information Retrieval.

2 Geschichte der Datenspeicherung

In diesem Kapitel werde ich auf das grundlegende Bedürfnis der Menschheit eingehen Informationen an Nachkommende und Interessierte weiterzugeben. Starten werde ich hier in der Zeit, als Höhlenmalerei noch das Mittel dazu war. Hierbei beziehe ich mich vielfach aus Daten aus dem Internet. Vor allem die Seiten von [Wiki07] wurden für die Erarbeitung des Textes über die frühe Geschichte der Datensammlungen zu Hilfe gezogen. Danach werde ich auf die Erfindung des Buchdruckes eingehen, wobei ich dafür auch wieder auf dieselbe Quelle zugreife.

Zu guter Letzt möchte ich, die aus heutiger Zeit als Geschichte geltenden Datenmengen, hervorgebracht durch unser digitales Schaffen, anhand von [HaLy00], [MoSe99], und im Internet mit Hilfe von [Dave96], [Isoc07], [Modi07], [Wiki07], [Yaho07] und [Cern90] aufzeigen. Den Abschnitt Internet heute konnte ich dank verschiedenen Quellen im Internet aufarbeiten, wie [BeHL01], [Mill07] und [Wire07].

2.1 Höhlenmalerei, Bücher und Informatik

Erste Ausprägungen des Menschen Daten an andere weiter zu geben kann man in der Höhlenmalerei entdecken. Die Bezeichnung Höhlenmalerei, Höhlenkunst oder Höhlenzeichnung charakterisiert bildliche oder grafische Illustrationen an den Wänden von Höhlen aus prähistorischer Zeit, vorwiegend der Altsteinzeit. So konnten aus diesen Kunstwerken wertvolle Anhaltspunkte über Kultur und Glauben der Steinzeitmenschen entnommen werden. Damals kannte man die Schrift noch nicht und trotzdem war der Wunsch nach Wissensvermittlung vorhanden.

Als mit den Schriftzeichen der Grundstein eines genaueren Weitergeben der kulturellen und intellektuellen Normen gelegt war und diese Normen mit der Erfindung des Buchdrucks in der Menge der Daten explosionsartig unter den Menschen verteilt werden konnte, wurden erste Möglichkeiten erfunden, Daten zu gliedern. So steigerten sich dank der Kennzeichnung durch Seitenzahlen³, Inhaltsverzeichnisse, Register sowie Titelblätter die Attraktivität und Strukturierung der Bücher. Durch die Erfindung und Etablierung des Letterndrucks vollzog sich eine bedeutende kulturhistorische Zäsur, die eine einschneidende Informationsverarbeitung eröffnete. Entsprechend fundamental angesehene Meilensteine gibt es nur wenige, wie zum Beispiel:

- Die Findung der Oralität und der Sprache,
- die Schöpfung der alphabetischen Schrift und der Literalität,
- das Wissen und die Etablierung des Buchdrucks sowie
- die Erfindung des Computers und die damit verbundene Digitalisierung.

So ist es also nicht verwunderlich, dass in der Medientheorie dementsprechend aus Gutenbergs Leistungen fundamentale Konsequenzen abgeleitet werden. So wurde durch die Erfindung des Buchdrucks der Faktor Mensch erstmals ersetzbar, denn während Bücher zuvor manuell in vorwiegend kirchlichen Skriptorien kopiert wurden, ermöglichte der Buchdruck die exakte Reproduktion von Wissen in einem zuvor nie gekannten Ausmass und es konnten zudem durch den Buchdruck lästige Abschreibfehler vermieden werden. So wurde die Autorschaft wichtig und es tauchten Fragen auf, wie, wann, wer, was, gesagt oder geschrieben und

³ Die Paginierung oder Seitennummerierung bezeichnet im Buchdruckwesen und in Akten und Handschriften das Nummerieren der Seiten eines mehrseitigen Schriftstückes.

präzise geäußert hatte. Somit wurde mit Problemen ähnlich dem heutigen Information Retrieval gekämpft, wie Metatags als Beschreibung von Daten Auskunft über die eigentlichen Daten geben.

Daneben veränderte sich auch das Lesen: Während Bücher zuvor laut (vor-) gelesen wurden, entwickelte man sich zum stillen Lesen. Das Denken in Anpassung an die Schriftform veränderte sich durch die Alphabetisierung und leitete eine tiefgründige Bildungsrevolution ein. Vielfach überboten die Methodik und die Wissenschaft neuer Bücher die als verständliche Form geltenden Metaphern. Bildung wurde umfassend zugänglich und man konnte Bücher ohne eine abstrakte Bildform verstehen, da nunmehr die gedruckten Bücher erschwinglicher als die handschriftlich reproduzierten waren, weil es jetzt etliche Exemplare eines Schriftstückes gab. Daher wurden als direkte Folge des Letterndruckes bekannte, bedeutungsvolle Schriften in den Umlauf gebracht.

Die Zeit von Gutenberg bis etwa 1965 wird häufig als „Bleizeit“ bezeichnet, danach folgte der bekannte Computersatz. Dies nehme ich als Grund um nach den einleitenden Worten chronologisch auf die letzten Jahren der Internetgeschichte einzugehen.

2.2 50 Jahre Internetgeschichte

Im folgenden Abschnitt möchte ich die letzten fünfzig Jahre des digitalen Zeitalters stark gekürzt Revue passieren lassen. Ich werde mich dabei auf die für den Kern der eigentlichen Arbeit, das Information Retrieval und die zugehörige Geschichte und Trends, wichtigen Ereignisse beschränken und die Daten wiedergeben, welche in meinen Augen als Pionierleistungen für ebendieses angesehen werden können.

Viele der gesammelten Daten konnte ich dem Buch über die Geschichte des Internets [HalLy00], dem Buch von Denis Moschitto und Everim Sen [MoSe99] und dem Internet [Dave96], [Modi07], [Wiki07] und [Cern90] selber entnehmen. Die erwähnten Daten sind in ein Geflecht, aus dem aus dem englischen übersetzten Berichten der Seiten von [Yaho07] und [Isoc07] eingebunden.

2.2.1 Die ersten Schritte

1957

Beginnen möchte ich genau vor fünfzig Jahren, als die UDSSR ihren ersten künstlichen Satelliten Sputnik ins All schickte und dadurch den Anstoss für die Erfindung des Internets einleitete.

Die Antwort der USA kann als Grundsteinlegung für unser heutiges Internet gesehen werden, denn die USA gründete die Advanced Research Projects Agency (ARPA). Diese Agency war im Departement of Defense (DoD) des Militärs angesiedelt. Das Ziel dieser Agency war dadurch definiert, die Spitze in der Forschung in technologischen Belangen in naher Zukunft übernehmen zu können, weswegen ARPA von Anfang an mit brillanten Köpfen der Zeit zusammenarbeitete.

1962

Paul Baran von der Rand Corporation wird von der U.S. Air Force beauftragt, ein Netz herzustellen, welches im Falle eines nuklearen Bombenangriffs noch funktionsfähig sein sollte, damit der Air Force noch die Möglichkeit bliebe, entsprechende Gegenmassnahmen zu koordinieren. Baran zeigte dafür das Konzept des packet switched network auf, welches heute

noch gebraucht wird und welches, bei Unterbruch von Verbindungen ein Wechsel auf andere, stabile Verbindungen zulässt.

Paul Baran schrieb damals in seinem Schlussdokument: *"Packet switching is the breaking down of data into datagrams or packets that are labeled to indicate the origin and the destination of the information and the forwarding of these packets from one computer to another computer until the information arrives at its final destination computer. This was crucial to the realization of a computer network. If packets are lost at any given point, the message can be resent by the originator."* [Dave96].

1968/1969

Die Forscher von Bolt Beranek and Newmann (BBN) konstruierten ARPANET, wessen physikalisches Netzwerk im Jahre 1969 vier Knotenpunkte verband. Die vier Knotenpunkte waren die University of California in Los Angeles, die SRI in Stanford, die University of California in Santa Barbara, und die University of Utah. Diese Punkte wurden damals durch ein 50 Kbps Kabel auf Basis eines Honeywell Minicomputer miteinander verbunden.

2.2.2 Die siebziger und achtziger Jahre

1972

Im Jahre 1972 wurde das erste Email Programm von Ray Tomlinson bei BBN entwickelt und im selben Jahr erfolgt die Umbenennung der ARPA in Defense Advanced Research Projects Agency (DARPA). Zu dieser Zeit verwendete ARPA noch das Network Control Protocol (NCP) um Daten zu transferieren, welches eine Kommunikation zwischen Hosts desselben Netzwerkes erlaubte.

1973

In diesem Jahr wurde mit der Entwicklung des später unter dem Namen TCP/IP bekannt gewordenen Protokolls durch Vinton Cerf und Bob Khan begonnen. Dieses Protokoll erlaubt es verschiedene Computernetzwerke miteinander zu verbinden und in diesen Netzwerken miteinander zu kommunizieren.

1974

Erstmalige Erwähnung des heute in jeder Sprache geläufigen Wortes "Internet" durch Cerf und Kahn in einem Papier zu TCP.

1976

Dr. Robert Metcalfe entwickelt Ethernet, welches auf der Grundlage von Koaxialkabel eine extrem schnelle Datenübertragung erlaubt. Dies war ein wichtiger Schritt für die spätere Entwicklung von Local Area Networks (LANs).

Ebenfalls in diesem Jahr wurde erstmals eine Satellitenverbindung umgesetzt, welche eine Verbindung von Europa und Amerika durch SATNET erlaubte. Überraschenderweise wurde diese Verbindung, INTELSAT, von einem Konsortium verschiedener Länder entwickelt und nicht, wie das meiste jener Zeit, unter der Leitung einer US Regierungsbehörde.

Das DoD experimentierte in diesem Jahr mit dem von Cerf und Kahn beschriebenen TCP/IP-Protokoll und entschied dieses für ARPANET einzusetzen.

1979

Durch die Erfindung von BITNET⁴ bei IBM wurden store and forward Netzwerke eingeführt, welche man vor dem Internet für den Emailverkehr brauchte. Die erste BITNET-Verbindung wurde 1981 zwischen der City University of New York (CUNY) und der Yale University hergestellt.

So gesehen war also das BITNET ein kooperatives Rechnernetzwerk, welches die Grossrechner von wissenschaftlichen Institutionen und öffentlichen Forschungseinrichtungen in den USA verband.

1981

Die National Science Foundation (NSF) kreiert CSNET, ein Netzwerk für Institutionen, jedoch ohne Zugriff auf ARPANET.

Im selben Jahr präsentiert Vinton Cerf einen Plan für das Zusammenarbeiten dieser verschiedenen Netzwerke CSNET und ARPANET, um eine reibungslose Entwicklung zu erlauben.

1983

Ab dem ersten Januar 1983 mussten alle Verbindungen zu ARPANET über TCP/IP abgewickelt werden, welches somit zum Kernprotokoll im Internet wurde und das NCP-Protokoll vollständig ersetzte.

Die Universität Wisconsin kreiert ein erstes Domain Name System (DNS), welches die Weiterleitung von Paketen zu einem Domain Name erlaubte. Dies wurde durch eine Übersetzung mittels einer Server Datenbank gelöst, welche die zugehörige IP-Nummer gespeichert hatte. Dadurch wurde es für viele einfacher andere Server aufzurufen, weil man sich nicht mehr an komplizierte Nummern erinnern musste.

1984

Das ARPANET wurde in MILNET und ARPANET geteilt. MILNET wurde danach vom amerikanischen Militär verwendet und ARPANET wurde weiterhin vom DoD für die Forschung gebraucht.

1986

Die Internet Engineering Task Force (IETF) wurde als Hilfe für technische Unterstützung für verschiedene Vertragspartner von der DARPA, dem US Defense Data Network (DDN) und dem Internet Core Gateway System gegründet. Die Hilfe wurde auf einer Plattform als Forum zur Verfügung gestellt.

⁴ BITNET steht als Abkürzung für Because Its Time Network und benutzte ein einheitliches Kommunikationsverfahren, das auch von Rechnernetzen wie EARN in Europa und NetNorth in Kanada eingesetzt wurde.

1987

Die zwei Netze BITNET und CSNET werden zusammengelegt und es entsteht ebenfalls in diesem Jahr die Corporation for Research and Educational Networking (CREN), eine neuen Gruppierung der NSF.

2.2.3 Die Schritte der Neunziger bis heute

1990

Merit, IBM und MCI gründen zusammen eine Non-Profit-Gesellschaft, genannt Advanced Network & Services (ANS), mit dem Ziel ein Hochgeschwindigkeitsnetz herzustellen. Die NSF integrierten das Netzwerk in ihr System und benutzten es Ende 1991 bereits als Backbone.

Tim Berners-Lee implementiert am CERN⁵ in Genf ein Hypertextsystem um Informationen für die international zusammenarbeitenden Physiker bereitzustellen.

1991

Auf CSNET wird das akademische Netzwerk in der ursprünglichen Form eingestellt. Weiter gibt das CREN neue Bedingungen heraus, dass operationale Kosten vollumfänglich durch die jeweiligen Benutzer getragen werden sollen.

Die NSF erstellt ein neues Netzwerk, genannt National Research and Education Network (NREN), mit dem Ziel für die Forscher ein schnelles, unabhängiges Netzwerk an die Hand zu geben, welches nicht für kommerzielle Zwecke gebraucht werden darf.

1992

Das bis heute verwendete, weltbekannte World Wide Web (WWW) wurde im Jahre 1992 durch das CERN (siehe dazu auch <http://info.cern.ch>) released.

1993

Unter der Federführung von InterNIC, ein von der NSF zur Verfügung gestellter Service, wurden ab diesem Jahr spezielle Dienstleistungen für die Registratur im Internet angeboten. Die aufgeteilten Dienstleistungen wurden durch AT&T (Verzeichnisse und Datenbanksysteme), durch Network Solutions Inc. (Registrations-Servicleistungen) und durch General Atomics/CERFnet (Informationsservice) zur Verfügung gestellt.

1994

Im Jahr 1994 wurden keine grossen Änderungen am physischen Netz vorgenommen und dennoch gab es eine grosse Veränderung im Internet: Es wuchs enorm! Dem NSF Backbone wurden viele neue Netzwerke angehängt und tausende von neuen Hosts wurden während dieser Zeit aufgeschaltet. Wie zum Beispiel Pizza Hut, welche erstmals Pizzabestellungen über ihre Webpage anboten. Ausserdem erwähnenswert ist die Eröffnung der ersten Cyberbank, der „First Virtual Bank“.

⁵ CERN steht für Organisation Européenne pour la Recherche Nucléaire, vormalis Conseil Européen pour la Recherche Nucléaire.

1995

Durch das enorme Wachstum forciert kündigte NSF an, dass ab dem 30. April 1995 keine direkten Verbindungen mit ihrem Backbone mehr hergestellt werden könne. Für dieses Vorhaben hat die NSF mit vier Firmen einen Vertrag ausgehandelt, welche als Provider auftreten sollten. Durch diese Provider konnten dann Verbindungen zum Internet gekauft werden. Für jährlich \$50 konnte man eine Domain reservieren, ausgenommen Domains mit der Endung .edu und .gov, welche immer noch für NSF reserviert waren.

1996

Dieses Jahr wurde durch den Anfang des hart geführten Browserkriegs zwischen den Konkurrenten Netscape und Microsoft geprägt, welcher vierteljährliche Releases forderte. Diese konnten vom jeweiligen Konsumenten als Beta Versionen heruntergeladen und ausprobiert werden.

1997

In diesem Jahr findet man erstaunliche 101'803 Name Servers Einträge in der "whois"-Datenbank⁶. Die „whois“-Datenbank ist eigentlich ein de-facto Protokoll, mit dem von einem verteilten Datenbanksystem Informationen zu Internet-Domains und IP-Adressen und deren Eigentümern abgefragt werden können.

1998

Microsoft versuchte dem Internet Explorer (IE) durch eine Strategie der Produktbündelung mit dem Betriebssystem zum Durchbruch zu verhelfen, was von vielen als wettbewerbswidrig angesehen wurde.

Ferner setzte Microsoft die Firma Compaq durch Kündigung der Vertriebslizenz für Windows 95 erfolgreich unter Druck, da sie ihre PCs zunächst mit Netscape anstelle des IE ausgeliefert hatte.

Das amerikanische Department of Justice sah in der Produktebündelung und diesem Vorgehen einen Verstoß und Microsoft wurde wegen Anti-Trust Verletzungen angeklagt.

2002

In diesem Jahr endet der Browserkrieg zwischen Microsoft und Netscape und man muss sich spätestens jetzt mit dem Schluss abfinden, dass Microsoft und der zugehörige IE neu nun die Nummer eins sind.

2.3 Das Internet heute

Wie in der Eröffnung schon erwähnt, wurden im März 2007 die Webseiten von Google zum ersten Mal von mehr Usern besucht als diejenigen von Microsoft. Im Jahr 2007 wird vor allem das Internet von der breiten Masse gebraucht und es werden rege Beiträge von verschie-

⁶ Whois-Anfragen (engl. wer ist) erfreuen sich noch grosser Beliebtheit, nicht zuletzt aus Gründen der Aktualität bei Domain-Lookups.

denen Menschen, häufig unredigiert, im Internet veröffentlicht. Die vorangegangenen Entwicklungen sind die meisten abgeschlossen oder schon Geschichte. Alle warten auf Web 2.0 und das vom WWW-Erfinder Tim Berners Lee angekündigte Semantische Netz (unter anderem [BeHL01]).

Laut einer, in der Einleitung kurz erwähnten Studie, ist Google die wertvollste Technologiemarke. Damit hat der Konzern Microsoft vom ersten Platz verdrängt, wie die Marktforschungsgruppe Millward Brown Optimor [Mill07] ermittelt hat. IBM konnte sich auf dem dritten Platz halten, der Wert der Marke sank aber um sieben Prozent. Nokia konnte den Wert um neunzehn Prozent steigern, blieb aber auf dem vierten Platz. HP kletterte vom siebten auf den fünften, Apple vom neunten auf den sechsten Platz. Dahinter folgten Cisco, Intel und Oracle.

Um Google schneller zu machen, investiert der Suchgigant in ein eigenes Glasfasernetz und stellt dazu weltweit zig Datacenter auf. Laut Google-Chef Eric Schmidt ist momentan ein Paradigmenwechsel im Gange, wie er der Zeitschrift Wired⁷ erklärte:

„It's pretty clear that there's an architectural shift going on. These occur every 10 or 20 years. The previous architecture was a proprietary network with PC clients called client-server computing. With this new architecture you're always online; every device can see every application; and the applications are stored in the cloud. It means that your servers are professionally managed, so you can actually have a weekend and not spend all your time trying to manage your servers. It's like having banks manage your money rather than you managing your money. And the networks have become secure, and the computers have become fast enough that this is mechanically possible - it actually works.“ [WIRE07, 9.4.07].

Um den ganzen Datenverkehr zu bewältigen stellt Google weltweit dutzende riesiger Datacenter auf. Ein weiterer Grund für Google in Glasfasernetze zu investieren ist, die eigene Kontrolle über diese.

⁷ Wired ist ein amerikanisches Technologie-Magazin. Es wendet sich an Technikfreaks, ist eher elitär geschrieben, aber trotzdem für die Allgemeinheit verständlich.

3 Evolution und Funktionsweise verschiedener Systeme im IR

In diesem Kapitel werde ich nun auf das Information Retrieval eingehen. Beginnen möchte ich mit einem kurzen Überblick über die Geschichte der Datenwiedergewinnung anhand einiger ausgewählter Momente. Ich verwende dazu als Quellen die Bücher [HaLy00] und [MoSe99] und aus dem Internet [Cern90], [Deja97], [Goog07], [Turb07], [Wiki07].

Für den zweiten Teil dieses Kapitels verwende ich Informationen aus einem Buch [BaRi99], aus Wissenschaftliche Arbeiten [BeHL01], [Hami03] und [Side05] und aus dem Internet [Dire07], [Goog07], [Musi07], [Retr06], [Turb07] und [Yaho07], um auf die heutigen Trends einzugehen und diese zu strukturieren.

3.1 Ausgewählte Momente in der Geschichte des IR

3.1.1 Archie

Archie⁸ wurde im Jahre 1990 durch die Studenten Alan Emtage, Peter Deutsch und Bill Heelan geschaffen, welche zu diesem Zeitpunkt als Studenten an der McGill University School of Computer Science in Montreal eingeschrieben waren. Sie begannen mit der Programmierung von Suchrobotern, die das Internet methodisch durchforsteten um Informationen zu sammeln. So entstand im November 1990 Archie V1.0. Archie war eine Suchmaschine, die speziell für das Indexieren von FTP-Archiven entwickelt wurde und über Telnet abgefragt werden konnte.

Ziel der Studenten war für Archie eine zentrale Datenbank zu schaffen, in der Verzeichnisse und Dateien der verteilten und weit verbreiteten Anonymous-FTP-Server enthalten sein sollten. Auf diese Weise sollte die Suche nach Dateien und Programmen über die Datenbank stark erleichtert werden.

Ein grosser Nachteil von Archie war jedoch, dass es nur möglich war Inhalte zu finden, von denen mindestens einige Fragmente und/oder Teile bekannt waren, da sich die Suche auf die Datei- und Verzeichnisnamen beschränkte, die auf den entsprechenden FTP-Servern lagen. Zudem standen für den Suchstring nur acht Zeichen zur Verfügung, sodass eine befriedigende Archie-Suche recht genaue Suchparameter erforderte.

Archie spielte jedoch, wegen der recht komplexen Bedienung und des aufkommenden WWW nur eine sehr untergeordnete Rolle im Internet. So rückte Archie immer weiter in den Hintergrund und viele Entwickler wandten sich den WWW-Technologien zu und nutzten diese als Basis für ihre Suchmaschinen.

3.1.2 WAIS

Unter der technischen Leitung von Brewster Kahle wurde 1991 der Suchdienst Wide Area Information Server System (WAIS) entwickelt, durch welchen erstmals eine komplette Volltextsuche in unterschiedliche Datenbestände über eine einheitliche Nutzeroberfläche ermöglicht wurde.

Um diesen Dienst zu nutzen benötigt man einen entsprechenden WAIS Client, welcher mit den jeweiligen WAIS Servern in Verbindung treten konnte. In der WAIS Datenbank wurden, ähnlich der zentralen Datenbank von Archie, alle Dateien aufgelistet, die im Internet erreichbar waren, jedoch führten diese Server einzelstehende Datenbestände. Dies heisst, dass die

⁸ Der Name entstammt dem englischen Wort "archive" (archivieren).

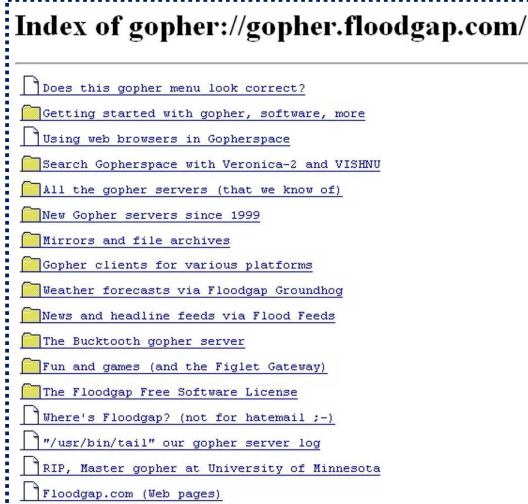
Datenbestände nicht zentral auf einer Datenbank gespeichert wurden und die WAIS Server keine Daten untereinander austauschten.

Auf jedem dieser Server waren ein oder mehrere Informationsangebote, wie zum Beispiel Geschichte, Chemie oder Medizin gespeichert. Für jede Datenbank gab es eine Source-Description, die Auskunft über den Inhalt der entsprechenden Datenbank sowie den dazugehörigen Server gab. Der Suchmechanismus von WAIS diente als Grundlage für weitere Entwicklungen wie Gopher und Veronica, welche im Folgenden kurz vorgestellt werden.

3.1.3 Gopher

Die umständliche Handhabung des File-Transfer-Protokolls (FTP), bei der man sich einloggen und über Konsolenbefehle die Verzeichnisse wechseln musste, um gewünschte Dateien zu finden und herunterladen zu können, führte 1991 an der Universität von Minnesota zur

Entwicklung von Gopher⁹. Gopher war ein einfach zu administrierendes Informationssystem, das wenige Ressourcen benötigte und mit dessen Hilfe man über das Internet mittels eines Gopher Client oder eines Webbrowsers entsprechend navigieren konnte.



Mitte der 90er Jahre hatte nahezu jede Organisation, welche über einen Internetzugang verfügte, einen Gopherserver und stellte Informationen aus verschiedensten Bereichen zur Verfügung. Um eine Gopherseite mit einem Browser aufzurufen, gibt man die Zeile `gopher://<gopherserver>/` ein. Eine Möglichkeit heutzutage den Gopherspace zu erforschen, bieten Webbrowser wie zum Beispiel Firefox und SeaMonkey¹⁰.

Abbildung 1: Index eines Gopherservers

Gopher bietet im Gegensatz zu HTML-Seiten ein automatisch generiertes Menü an, das aus den im aktuellen Verzeichnis befindlichen Dateien generiert wird. Der Gopher-Server erkennt dabei, ob es sich um Verzeichnisse oder Dateien handelt, und zeigt dies durch entsprechende Symbole an. Zusätzlich können auch Konfigurationsdateien eingebunden werden, die dem Betreiber erlauben, Verweise auf externe Gopherserver zu generieren. Im Gegensatz zu Webseiten sind Gopherseiten reine Textdateien ohne irgendwelche Formatierung.

Mit dem Aufschwung des WWW endete die Zeit des besagten Gopherspace. Ein wesentlicher Punkt für den Untergang von Gopher war die Erhebung von Gebühren für die Benutzung.

3.1.4 Veronica

Veronica, dessen Akronym für Very Easy Rodent-Oriented Netwide Index to Computerized Archives steht, wurde im Jahre 1992 an der Universität von Nevada in Reno entwickelt. Veronica ist ein menüorientierter Suchdienst für den Gopherspace, der bis etwa Mitte

⁹ Englisch für Erdhörnchen. Für die Wahl des Namens gibt es zwei Theorien: Erstens, das Maskottchen des Bundesstaates und der Universität von Minnesota, oder zweitens „Go for it“.

¹⁰ Bis 2002 unterstützte auch der Internet Explorer von Microsoft das Gopher-Protokoll. Danach wurde die Funktion wegen Sicherheitslücken im Programm deaktiviert.

der 1990er Jahre genutzt wurde und dann mit dem Untergang von Gopher durch das WWW und die eigenen Suchmaschinen abgelöst wurde.

Funktional leistete Veronica das für den Gopherspace, was der Suchdienst Archie für FTP tut. Veronica indizierte die Titel von Dateien und Verzeichnissen aller der University of Minnesota bekannten Gopherserver und weiter alle Gopherserver, auf die von diesen Servern aus verwiesen wird.

Veronica bietet als gravierenden Vorteil gegenüber der Suchmaschine Archie, dass die Titel sich nicht nur auf den Dateinamen beschränken, sondern auch aus ganzen Sätzen bestehen konnten. Zudem waren auch logische, boolesche Abfragen mit or-, and- und not- Verknüpfung der Suchbegriffe möglich.

3.1.5 Mosaic

1993 wurde rund um das Entwicklungsteam von Eric Bina und Marc Andreessen die Schöpfung des NCSA Mosaic gestartet und zwar als Projekt der Software Group am NCSA an der Universität von Illinois. Der NCSA Mosaic war der erste Internetbrowser¹¹, der ausser Text auch Grafiken auf einer Seite zeigte, ohne dass man diese extra laden musste. Die Darstellung der Graphiken erforderte die Formate GIF und HDF.

So wurde vom Viola-Browser, der zwar auch Grafiken anzeigen konnte, dafür jedoch erst auf einen entsprechenden Verweis im Text geklickt werden musste, die Schaltknöpfe für die Startseite und die Navigation übernommen und vom Textbrowser Lynx die Lesezeichen.



Abbildung 2: Mosaic Browser Emulation

Mit der zu der Zeit schnell wachsenden Popularität des WWW wurden Webbrowser schnell bekannt und somit begann der Name Mosaic zum Synonym für Webbrowser zu werden. Auch die ersten, von ehemaligen NCSA-Mitarbeitern geschriebenen Netscape-Browser, wurden Netscape Mosaic getauft, jedoch änderte man diese im Zuge der Kommerzialisierung des Browser-Marktes auf den Namen Netscape Navigator. Dieser sollte dann später selber zu einem Synonym für Webbrowser werden.

Anfangs Januar 1997 wurde die Entwicklung von NCSA Mosaic mit dem Erscheinen von Version 3.0 eingestellt.

3.2 Gegenwärtige Trends im IR

In diesem Kapitel möchte ich kurz die heutigen Trends vorstellen. Laut Baeza-Yates und Ribeiro-Neto [BaRi99, Seite 389ff] geben Suchmaschinen der heutigen Generation immer noch zu viel Heu mit der sprichwörtlichen Nadel zurück und Web Directories haben ihrer Meinung

¹¹ Beta-Versionen erschienen ab September 1993 für die Unix-Betriebssysteme SunOS, IRIX, AIX, OSF/1, HP-UX und mit X-Window-System, für Apples Macintosh System 7, sowie für Amiga, Windows NT auf DEC Alpha-, PowerPC- und MIPS-Rechnern. Versionen für x86-Hardware und Windows 3.1 und 95 folgten später.

nach zu wenig Tiefe, die Nadel zu finden. Es sei leichter, den Benutzer zu einer effektiven Suche zu bewegen, als den Suchmaschinen das Raten zu überlassen, was der Suchende wirklich will. Doch ist das wirklich noch so?

Wie es aussieht, hat die Entwicklung in den letzten Jahren keine grossartigen Sprünge mehr vollbracht. Auf meiner Suche nach neusten Erkenntnissen in diesem Bereich stiess ich mehrheitlich auf zwei Jahre alte Daten. Diese Berichte aus der Forschung zeigen die Schwierigkeiten auf, mit welchen man heute im Bereich der Informationssuche konfrontiert ist. Allerdings kann man alleine schon durch die Nutzung von Google feststellen, dass viele Ideen vorhanden sind, denn die neusten Errungenschaften werden da direkt angekündigt (wie zum Beispiel <http://labs.google.ch/>) oder diese wurden schon umgesetzt (wie zum Beispiel <http://www.google.ch/intl/de/options/>).

Deshalb möchte ich, nicht abschliessend allerdings, einen Überblick in die heutigen Forschungen geben. Ich tue dies mit einer Eingliederung der Forschungsergebnisse und –Gebiete in vier Kategorien; die Informationsressourcen, die spezialisierten Suchmaschinen, die Personalisierung und die Lokalisierung.

3.2.1 Informationsressourcen

Um Informationen, welche in der Tiefe des Invisiblen Webs verborgen sind sichtbar zu machen, werden verschiedene Versuche unternommen diese Inhalte des Invisible Webs zu erfassen. Die wichtigsten Ansätze hierzu sind, erstens der Versuch den Bestand des Invisible Webs in das Surface Web zu holen, zweitens der Weg, welcher von Google und Yahoo im Moment gegangen wird, über die Auswahl durch Registerkarten und drittens die Integration des Invisible Webs direkt in die Suche.

Zum ersten Weg, des Versuchs den Bestand des Invisiblen Netzes an die Oberfläche zu holen und nutzbar zu machen, gibt es den Weg Datenbankinhalte direkt in HTML anzuzeigen oder die Invisible-Web-Quellen direkt im Meta-Suchinterfaces zu integrieren, wie dies zum Beispiel die Metasuchmaschine Turbo10 tut. Diese Meta-Suchmaschine erkennt Datenbank-Suchmasken und leitet Anfragen automatisch an die entsprechende Datenbank weiter. Der Benutzer kann seine Suchanfrage direkt an die ausgewählten Datenbanken schicken. Am Schluss erfolgt, wie bei einer gewöhnlichen Suchmaschine, die Auflistung der Suchergebnisse nach einem Ranking.

Bei der zweiten Variante, des Sichtbarmachens von bisher ungenutzten Daten, wird der Weg über Register gegangen. Man kann hierzu spezielle Suchen zum Beispiel von Bildern, Groups oder neu auch Blogs, usw. starten (vergleiche hierzu <http://www.google.ch/intl/de/options/>). Zudem wurde unlängst in Google und eher noch bei Yahoo, wie als dritte Möglichkeit erwähnt, die Funktion integriert durch entsprechende Eingaben direkt an die entsprechende Datenbank weitergeleitet zu werden. Z.B. wird man bei Yahoo bei der Eingabe von Wetter plus zugehöriger Stadt direkt an eine Wetterdienstdatenbank weitergeleitet oder als weiteres Beispiel bei wird bei Google nach der Eingabe zweier Städte der Fahrplan zwischen diesen Städten ausgegeben.

3.2.2 Spezialisierte Suchmaschinen

Im Moment werden wegen der Unzulänglichkeit der Suchmaschinen Unterteilungen vorgenommen um dem Benutzer ein gezieltes Suchen in gewünschten Datenbeständen zu ermöglichen. Wenn man sich zum Beispiel mit dem Music Information Retrieval (<http://www.musipedia.org/>) beschäftigt, wird eine andere, spezialisierte Suchmaschine be-

vorzugt, als wenn man beispielsweise nach Bildern (<http://labs.systemone.at/retrievr/>) sucht. Diese Suchmaschinen haben sich auf spezielle Themen beschränkt.

Man kann hierzu eine Unterscheidung nach formalen Kriterien, wie fachlich (nach Fachgruppe aufgeteilt), zeitlich (Echtzeit vs. Archive) und/oder geographisch (Regionen, Sprachen, Länder, usw.) heranziehen.

Weiter können Unterscheidungen nach inhaltlichen Kriterien gemacht werden, wie zum Beispiel Nachschlagewerke (Personen, Adressen, Wörterbücher, usw., wie dies auf der Seite von Swisscom www.directories.ch getan wird), Nicht-Text-Retrieval (wie oben als Beispiel genannt Music IR, Content Based IR, usw.), Literaturtypen (Zeitschriften, Hochschulnetze, usw.), Blogs, usw.

Eine Einteilung ist diesbezüglich bis anhin noch wichtig und es ist noch nicht möglich, alles mit einer einzigen Suchmaschine abzudecken.

3.2.3 Personalisierung

Google machte dieses Jahr in verschiedenen Printmedien Schlagzeilen, als sie den Dienst der personalisierten Suche aufschalteten. Doch was versteckt sich hinter der personalisierten Suche?

Bei der personalisierten Suche geht es darum, dass Daten von bisher getätigten Suchen in Cookies gespeichert werden, um bei einer erneuten Suche als Kriterium miteinbezogen zu werden.

Folgendes Beispiel soll die Berücksichtigung früher gesuchter Daten verdeutlichen: Ein Autohändler sucht neue Autos. Er tippt in seine Suchmaske Jaguar ein und erhält dazu Ergebnisse vom Typ Auto und nicht vom Tier. Dies kann durch das System anhand bereits gesuchter Daten unterschieden werden. Eine weitere Möglichkeit ist die Suche nach Hotels: Hier kann anhand der Personalisierung unterschieden werden, ob ein Hotel für einen Manger und Meetings gesucht werden soll, oder aber ein Hotel für Ferien (vergleiche hierzu auch [Goog07] und <http://labs.google.ch/>).

Für die Personalisierung wurde bis anhin eine manuelle Einstellung vorgenommen, aber durch das neue Verfahren mit welchem Google solche Schlagzeilen generierte, werden nach einem Anmelden bei Google diese Daten auf den erwähnten, neuen Servern von Google gespeichert und als Kriterium miteinbezogen. Die Kritik der Medien war alle auf den Datenschutz hin gerichtet.

3.2.4 Lokalisierte Suche

Um Anbieter aus der Umgebung zu suchen, wird es notwendig eine lokal beschränkte Suche starten zu können. Dies, weil man vorwiegend Dienstleistungen und Angebote aus der näheren Umgebung suchen möchte, sei dies, weil man die Gebräuche und Gesetze dieser Region kennt und auf diese vertraut, oder aber aus Gründen von Transport, usw.

Leider führen solche Suchanfragen bei den Suchmaschinen oft noch nicht zu den ersehnten Ergebnissen, trotz ergänzenden Ortsangaben. Selbst wenn relevante Ergebnisse gefunden werden, ist die Vollständigkeit der Ergebnisse nicht garantiert. Für diesen Zweck ist es vielfach sinnvoll auf lokal agierende Suchmaschinen auszuweichen (wie z.B. www.directories.ch). Hier bekommt man einen Überblick über die einzelnen Anbieter einer Region.

In Europa kann man allerdings noch nicht mit den grossen Suchmaschinen auf eine vergleichbare Art und Weise suchen. Anders in den USA, da haben inzwischen alle grossen Suchmaschinen eigene lokale Suchen eingerichtet. So bieten Yahoo (www.maps.yahoo.com) und Google (www.local.google.com) lokale Suchmöglichkeiten an.

4 Fazit / Schlussfolgerungen

An dieser Stelle möchte ich ein Fazit geben und eine kritische Hinterfragung bezüglich der Trends im Information Retrieval auflisten. Wichtig ist mir klarzustellen, dass dies meine persönliche Meinung ist und es jedermann frei steht, eine andere Meinung zu vertreten. Meine Gedanken können ausserdem als Grundlage für eine weiterführende Diskussion genommen werden.

Ich denke, in Zukunft müssen sich die grossen Anbieter von Suchdienstleistungen neben Google verbessern, um überhaupt bestehen zu können. Als weiteren, nicht minderen Grund, müssen die Suchdienstleister sich verbessern um dem Suchenden einen Mehrwert zu bieten. Dies könnte durch eine bessere, ergonomische, auf den jeweiligen Sucher angepasste Suche sein (individuelle Suchmaske, optimierte, an den Bewegungsablauf angepasste Eingabe, usw.). Wünschenswert wäre auch das Miteinbeziehen der im Moment noch spezialisierten Suchen auf eine einzige Plattform, wobei der erwähnte ergonomische Nutzen nicht ausser Acht gelassen werden darf.

Anhand der folgenden Kriterien sieht man, dass momentan ein Trend der Strukturierung und Gliederung im Netz stattfindet, welcher den Suchmaschinenanbieter in die Hände spielt. Meiner Meinung nach wird es in Zukunft wieder einen Trend hin zu redigierten Informationen und Daten geben, denn nachdem sich nun jeder über seinen Blog der Welt mitgeteilt hat, verliert dies an Attraktivität. Weiter gibt es in genau diesem Bereich immer mehr professionelle Blogger, wie z.B. in der Schweiz als prominentes Vorbild Bundesrat Leuenberger (<http://moritzleuenberger.blueblog.ch/>). Dadurch dass dieser sicherlich seine Informationen professionell aufarbeiten lässt (schon aus Gründen des Datenschutzes) wird meine Theorie hin zu redigierten Daten gestützt. Zudem warten alle auf das Semantische Web und/oder Web 2.0, welches die Strukturierung der Daten besser unterstützen soll.

Abschliessen möchte ich mit einer Vision für die Zukunft der Datensuche. Wünschenswert, falls überhaupt möglich, wäre ein System, welches schon im Voraus Daten zur Verfügung stellen kann, ohne dass extra gesucht werden muss. Diese Daten werden bei Bedarf angeschaut, oder eben auch nicht. Zudem müssten die entsprechenden Daten natürlich den theoretisch heute schon geforderten Gegebenheiten entsprechen, wie die Integration der vielen spezialisierten Suchen auf einer Plattform. Oder aber die übergeordnete Suche müsste in der Lage sein, die richtigen Untersuchmaschinen aufzurufen. Unter anderem könnte dies eventuell mit Hilfe der häufig gelobten Künstlichen Intelligenz möglich werden.

So hoffe ich abschliessend, dass die Zukunft uns viel Spannendes und Brauchbares in diesem Bereich bringen wird. Hoffentlich wird die Zukunft rund um das Internet genau so spannend wie die Geschichte der letzten 50 Jahre.

5 Abbildungsverzeichnis

Abbildung 1: Index eines Gopherservers	11
Abbildung 2: Mosaic Browser Emulation	12

6 Literatur

6.1 Bücher

- [BaRi99] Baeza-Yates R.; Ribeiro-Neto B.: Modern Information Retrieval. Verlag ACM press, Essex 1999
- [HaLy00] Hafner K.; Lyon M.: Die Geschichte des Internet. 2., korrigierte Auflage. Verlag dpunkt.verlag gmbH, Heidelberg 2000
- [MoSe99] Moschitto D.; Sen E.: Hackerland. Das Logbuch der Szene. Verlag Tropenverlag, Köln, 1999

6.2 Wissenschaftliche Arbeiten

- [BeHL01] Berners-Lee T.; Hendler J. und Lassila O.: The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, Scientific American, 2001
- [Hami03] Nigel Hamilton. The Mechanics of a Deep Net Metasearch Engine. Online, Internet, gesehen am 13.07.2007: <http://turbo10.com/papers/deepnet.pdf> 2007
- [Modi07] Mike Modiset. Consumers and their quest to control the internet and advertising. Online, Internet, gesehen am 14.07.2007: http://www.ciadvertising.org/sa/spring_07/adv391k/mdm377/midterm/history.html 2007
- [Side05] Virginija Siederkeviciute. Seminar Angewandtes Information Retrieval. Trends im Information Retrieval. Abteilung Geisteswissenschaftliche Informatik, Universität Basel 2005

6.3 Internet

- [Appl07] Apple. Online, Internet, gesehen am 16.07.2007: <http://www.apple.com/> 2007
- [Cern90] CERN. Online, Internet, gesehen am 18.05.2007: <http://info.cern.ch> 1990
- [Cisc07] Cisco. Online, Internet, gesehen am 16.07.2007: <http://www.cisco.com/> 2007
- [Come07] ComScore . Measuring the Digital World. Online, Internet, gesehen am 18.05.2007: <http://www.ComScore.com/press/release.asp?press=988> 2007
- [Dave96] Dave's Site. www.davesite.com. Saving the World Wide Web since 1996. Online, Internet, gesehen am 18.05.2007: <http://www.davesite.com/webstation/net-history.shtml> 1996
- [Deja97] Dejavu. Browser Emulator. Online, Internet, gesehen am 18.05.2007: <http://www.dejavu.org/emulator.htm> 1997-2000
- [Dire07] Directories. Die Datenquelle. Online, Internet, gesehen am 17.07.2007: <http://www.directories.ch/>
- [Ebay07] eBay. Online, Internet, gesehen am 16.07.2007: www.ebay.com 2007
- [Goog07] Google. Online, Internet, gesehen am 16.07.2007: www.google.com 2007

- [Hewl07] Hewlett-Packard. Online, Internet, gesehen am 16.07.2007:
<http://www.hewlett-packard.com/> 2007
- [Ibm07] IBM. Online, Internet, gesehen am 16.07.2007: <http://www.ibm.com/us/> 2007
- [Inte07] Intel. Leap ahead™. Online, Internet, gesehen am 16.07.2007:
<http://www.intel.com/> 2007
- [Isoc07] Internet Society. Online, Internet, gesehen am 18.05.2007:
<http://www.isoc.org/internet/history/brief.shtml> 2007
- [Micr07] Microsoft. Online, Internet, gesehen am 16.07.2007: www.microsoft.com 2007
- [Mill07] Millward Brown. Online, Internet, gesehen am 16.07.2007:
<http://www.millwardbrown.com/Sites/MillwardBrown/Content/News/PressReleases.aspx> 2007
- [Musi07] Musipedia. The Open Music Encyclopedia. Online, Internet, gesehen am 16.07.2007: <http://www.musipedia.org/> 2007
- [Noki07] Nokia. Connecting People. Online, Internet, gesehen am 16.07.2007:
<http://www.nokia.com/> 2007
- [Orac07] Oracle. Online, Internet, gesehen am 16.07.2007:
<http://www.oracle.com/index.html> 2007
- [Retr06] Retrievr. Online, Internet, gesehen am 17.07.2007:
<http://labs.systemone.at/retrievr/> 2006
- [Time07] TimeWarner. Online, Internet, gesehen am 16.07.2007: www.timewarner.com 2007
- [Turb07] Turbo10. Search the Deep Net™. Online, Internet, gesehen am 13.07.2007:
www.turbo10.com 2007
- [Wiki07] Wikipedia. The Free Encyclopedia. Online, Internet, gesehen am 16.07.2007:
www.wikipedia.org 2007
- [Wire07] Wired. Text of Wired's Interview with Google CEO Eric Schmidt. Online, Internet, gesehen am 18.05.2007:
http://www.wired.com/techbiz/people/news/2007/04/mag_schmidt_trans?currentPage=all 2007
- [Yaho07] Yahoo. Online, Internet, gesehen am 16.07.2007: www.yahoo.com 2007
- [Yout07] YouTube. Broadcast Yourself™. Online, Internet, gesehen am 16.07.2007:
www.youtube.com 2007